The Weight of User Decision Making During Online Interactions - Planning an Experiment

Chris Barry *chris.barry@nuigalway.ie*
National University of Ireland Galway
Galway, Ireland

Mairéad Hogan *mairead.hogan@nuigalway.ie*
National University of Ireland Galway
Galway, Ireland

Abstract

This paper lays out the design of a research study, using eye tracking technology, to measure participant cognitive load when encountering decision constructs during webpage interactions. It elaborates and improves on a pilot study that was used to test the experiment design. Cognitive load is discussed in detail, in both physiological and subjective terms, as well as techniques to capture participants' thoughts and feelings immediately after the experiment. This mixed method approach will generate a more holistic comprehension of participants' decision making and their rationale; and hopefully, improve information systems design ethics.
Keywords: Eye tracking, decision constructs, cognitive load, NASA-TLX, ethics in ISD.

1. Introduction

The impetus for this research came from observations that some firms seemed to be making life rather difficult for users engaged in an online transactional process such as pricing a flight, finding contact information or making a complaint. Early research [3], [56] established through heuristic evaluations, verbal protocols, an extensive usability study and focus groups indicate that not alone did these phenomena exist but that some firms were clearly in violation of compliance with EU legislative provisions. Examples of issues that caused confusion and irritation amongst users include: finding the 'real' price; understanding taxes and charges; avoiding optional extras and understanding the nature of decisions that confront them.

This paper describes a research experiment designed to measure, both quantitatively and qualitatively, the cognitive load on participants' making decisions on a series of webpages. Several methods will be used to gather data: eye tracking equipment to gather interaction data; a scale to measure subjective cognitive load and a think-aloud technique to gather thoughts and opinions. This study builds on a pilot [27] conducted to confirm the experiment design and identify flaws and improvements to be made in the process.

1.1. Types of Decision Constructs

From earlier research into the B2C commercial transactional process, seven distinct decision constructs encountered by users were identified [28]. Two of these are essential decision constructs, while five are optional. Essential decisions relate to decisions that must be made to conclude a transaction (e.g. choosing a delivery method or selecting a shirt size). In the 'classical' case, optional constructs do not require user interaction – they are options. However, the classical case is an endangered species and in fact, there are a variety of ways that optional decisions may be presented to users that necessitate pause, reflection and interaction. Collectively, these decision constructs form a taxonomy of discrete, mutually exclusive types (see Table 1).

Desk analysis of fifty-seven websites was conducted to identify the various dimensions of option presentation of each decision type within the taxonomy [3]. Three dimensions were considered to be of critical importance:

- default value (i.e., un-selected or pre-selected);
- question or information framing (i.e., acceptance, rejection or neutral language);
- additional persuaders (e.g., benefits of choosing the option, risks of rejecting the option, reassurance of privacy).

The nature of the interaction of decision types revealed that opt-outs and must-opts were highly problematic for users to navigate because the option presentation lacked clarity. Sometimes this was in the choice of language and other times it was the fundamental nature of the decision construct. It should be noted that a pre-selected opt-in is also a problematic construct, but is inherently illogical. It is most likely to be encountered because of poor design practice rather deliberate intent.

Table 1. Taxonomy of transactional decision constructs.

Decision Types	**Description**
Un-selected opt-in	• Default: don't receive the option • Normal presentation: un-selected • Framing: acceptance
Pre-selected opt-in	• Default: don't receive the option • Normal presentation: selected • Framing: rejection
Un-selected opt-out	• **Default: receive the option** • **Normal presentation: un-selected** • **Framing: rejection**
Pre-selected opt-out	• **Default: receive the option** • **Normal presentation: selected** • **Framing: acceptance, rejection or neutral**
Must-opt	• Default: cannot proceed • Normal presentation: multiple option variants, one of which allows the option to be declined, all variants un-selected • Framing: acceptance, rejection and/or neutral
Un-selected essential decision	• Default: cannot proceed • Normal presentation: multiple decision variants, all un-selected • Framing: acceptance, rejection and/or neutral
Pre-selected essential decision	• Default: variant selected • Normal presentation: multiple decision variants, one selected • Framing: acceptance, rejection and/or neutral

In a pilot study of the opt-out decision constructs, the dimensions of default value, framing and persuasion were tested using eye tracking technology (see highlighted decision types in Table 1). The analysis indicates framing to be the most influential dimension in the interaction [28]. Decision framing is important and designers may present it in acceptance, rejection, or neutral terms (see Figure 1). Incidences of each presentation framing type were identified. Default value and persuasion appear to be less influential on participants' decision making. In particular, persuasion is not a binary factor; it forms a continuum which has a scale that is variable and non-specific. It solicits a subjective response from respondents that is difficult, arguably not possible, to categorise. Therefore, for the current study, only framing and default value will be considered.

1.2. Cognitive Load

According to Grimes and Valacich [22] cognitive load, or mental workload, can be defined as: "the mental effort and working memory required to complete a task". The negative impact of cognitive load on learning is frequently considered in the development of computer based learning applications [5], [29], [34], [39]. Its relevance to e-commerce transactions is less researched. However, higher cognitive loads have been shown to negatively affect time to complete a task and user satisfaction in e-commerce applications [51]. Additionally, higher mental workload corresponds to lower perceived usability for webpages [37].

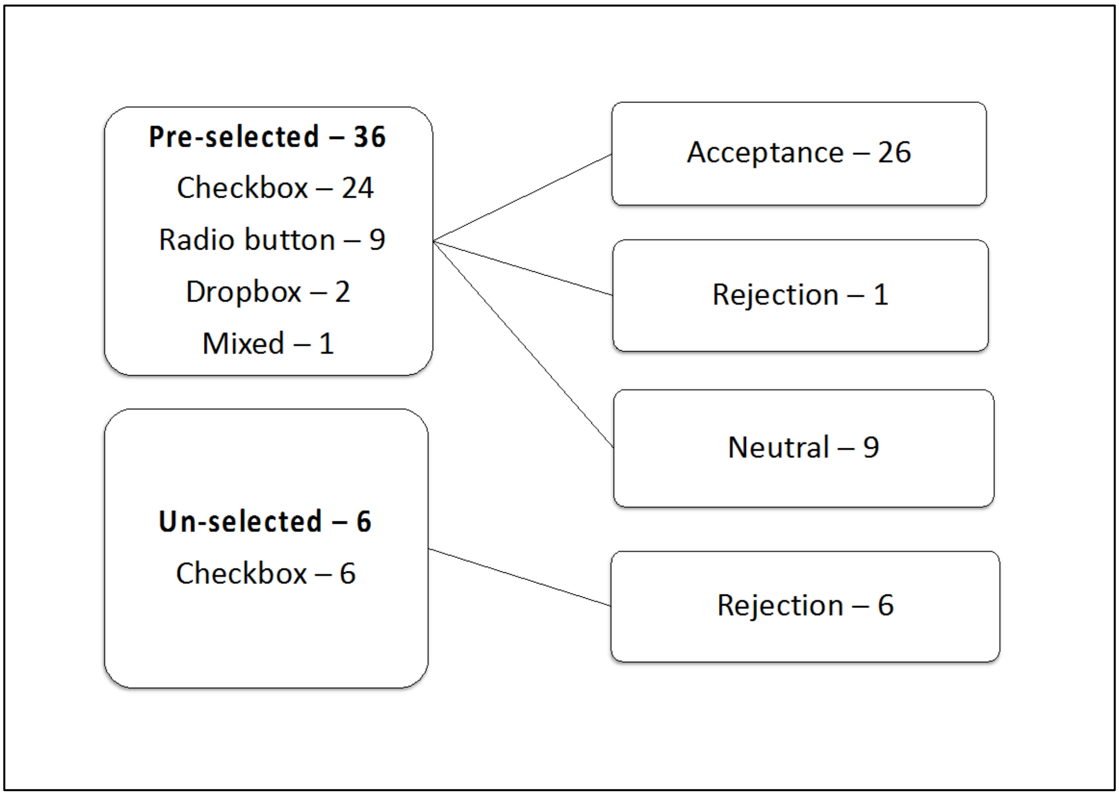

Fig. 1. Incidences of opt-out framing identified in desk analysis.

Cognitive load can be measured in several different ways. The main approaches include: subjective measures; direct objective (or physiological) measures; and indirect objective measures (for example, electroencephalography (EEG) or cardiovascular metrics) [39]. The subjective measures generally use Likert type scales for self-reporting of stress or other indicators of mental load. Some of the more commonly used measures include the Subjective Workload Assessment Technique (SWAT), the NASA-Task Load Index (NASA-TLX) and the Workload Profile (WP). Each of these measures lead to a global workload index that is sensitive to the level of difficulty in the task [49]. Think-aloud can also be used to measure cognitive load [13], albeit qualitatively, rather than quantitatively.

The direct objective measures include eye tracking for the measurement of eye movement, pupil dilation and blink rate; and dual-task methodologies [39]. Horizontal Eye Movements (HEM) increase as mental workload increases [23]. Further pupil size is correlated to cognitive load in physicians [55], while a reduction in blink frequency and shorter blink duration correlated with higher levels of self-reported cognitive load in a study of surgeons [60]. While these measures are sensitive to cognitive load, they can also be impacted by other factors, such as fatigue [6].

1.3. Subjective Measurement Scales

Measurement scales commonly used to determine cognitive load include uni-dimensional scales, such as the Modified Cooper-Harper Scale (MCH), and the Overall Workload Scale (OW), as well as multi-dimensional scales, such as NASA-TLX and SWAT [40]. Rating scales require the user to indicate the mental effort required to complete a task. Research indicates people can put a numerical value on their perceived mental effort [21], [46], resulting in their use in much research. NASA-TLX and SWAT are the most commonly used measurement scales of subjective cognitive load [15]. However, SWAT is not sensitive for low cognitive load tasks, unlike the NASA-TLX [38]. Hence, NASA-TLX is considered to be superior to SWAT in terms of sensitivity [27] and is frequently used as a benchmark when assessing other measures [1], [14], [60].

The use of NASA-TLX has spread extensively beyond its initial aviation domain, including the medical profession, data entry and decision making and has been translated into multiple different languages [24]. Hart [24] examined 550 studies in which NASA-TLX was used and found most of these studies were concerned with some form of question relating to interface design or evaluation. The scale had been modified in many of the studies examined, with subscales being added, deleted or modified. Many studies also modified the method by eliminating the weighting. Others analysed the subscales individually, either in conjunction with, or instead of, the overall workload measure. Hart [24] concluded “NASA-TLX has achieved a certain venerability; it is being used as a benchmark against which the efficacy of other measures, theories, or models are judged”.

1.4. Eye Tracking

Eye-tracking technology operates on the basis of focusing a light and a video camera on a person’s eye to determine where an individual is looking on a screen [44]. When people pay attention to something, they fix their gaze on it and it comes into sharp focus. This focus is referred to as a fixation. These fixations are recorded by the eye tracking software. In the normal course, a person typically moves their eye across various items of interest on the screen. These movements are referred to as saccades. Typically, they are jerky and happen so quickly we are not aware of them. The saccades last between one-hundredth and one-tenth of a second. They are, in common terms, flicking around the screen looking for something of interest. Fixations, which last between one-tenth and a half second [44], are more deterministic in nature and, while brief, have engaged the user in some form of interaction.

Eye-tracking has been used extensively in web usability studies [10-12], [32], [53] and is frequently used in conjunction with think-aloud techniques. A comparison study of a number of usability testing techniques for an e-commerce website found the use of Retrospective Think-aloud with Eye movement (RTE) identified more usability problems than Retrospective Think-Aloud (RTA), observation or Feedback Capture After Task (FCAT) [17]. By studying what users look at in an interaction it is possible to determine where they are concentrating their attention [47]. This approach offers powerful insights to the decision making strategies individuals employ when involved in what would otherwise be an invisible interaction. Through the examination of eye movement patterns, conclusions can be made about fundamental aspects of the user’s online activity [9], [16], [32].

2. Research Approach

The pilot study mentioned earlier [4], examined how users interact with differently designed opt-out decision constructs. It consisted of two parts: tracking user eye movement while interacting with differently designed decision constructs; and cued Retrospective Think-Aloud (cued RTA) sessions where the participant articulated their thoughts and feelings regarding the interaction after completion. This proposed study builds on the pilot. In addition to the eye tracking and the cued RTA, the study will also uses a variety of measures to determine the cognitive load on the participant during the interaction. The ways in which this is measured will be three-fold: physiological measures, subjective measures using a subset of the NASA-TLX and attitudinal data identified during the cued RTA sessions. The essential research question is: to what extent does decision construct design positively or negatively affect cognitive load during webpage interactions?

2.1. Measuring Cognitive Load

Physiological Cognitive Load

Physiological measures are often used to determine cognitive load, based on the assumption that the individual's physiological system will react as the cognitive demands of the task change [50]. For this study, several eye tracking metrics (blink rate, dwell time and fixation count) will be used to determine participants' reaction to stimuli [18], [39], [52]. Dwell time measures the duration of fixations within a particular area of interest (AOI), while fixation count specifies the number of fixations on an AOI. Blink rate varies with cognitive load, depending on the task. It declines for high cognitive load when processing visual stimuli but increases for high cognitive load in tasks involving memory [6]. Dwell time [6], [36], [39], [48] and fixation count [36], [39], [48] have also been shown to increase with cognitive load. It is generally recommended that a single eye tracking measure is not used in isolation [7], [39], [50] as they can be influenced by factors such as fatigue [6] and age [39]. Therefore, for this study, both primary (dwell time, fixation count) and secondary (blink rate) measures [7], as well as both qualitative and quantitative subjective measures of cognitive load, will be used.

Subjective Cognitive Load

NASA-TLX

Based on the research detailed above (see Section 1.3), NASA-TLX is the most appropriate measurement scale to use in this study. NASA-TLX was the culmination of a multi-year research programme that resulted in a multi-dimensional rating scale, and derives an estimate of workload that is both reliable and sensitive [25]. The programme determined the contributing factors to an individual's subjective perception of physical and mental workload. These were narrowed down to 6 factors or scales: mental demand; physical demand; temporal demand; performance; effort; and frustration level. The definitions for these can be seen in Table 2.

Table 2. Rating scale definitions for NASA-TLX.

Rating Scale Definitions	
Scale	**Definition**
Mental demand	The level of mental and perceptual activity required for the task
Physical demand	The level of physical activity required for the task
Temporal demand	The level of time pressure felt
Performance	The level of success in reaching the goals of the task
Effort	The level of work, both mental and physical, required
Frustration level	The level of frustration felt during the task

According to the NASA-TLX user manual [43], each of these factors are weighted by the participants according to their perception of the contribution of each factor to the workload of a given task. This weighting can be done while carrying out the task, or afterwards while replaying the task for the participant. The participant assigns a score on a 21-point scale ranging from 0-100 on each factor. The measure, as described in the NASA-TLX manual, also requires the participant to weight each of the factors by indicating which one was most relevant to the task in a series of paired comparisons. However, more recent studies [8], [22], [31], [59] have used a slightly modified version of the NASA-TLX, known as NASA-Raw Task Load Index (NASA-RTLX). Rather than weighting the factors, each is assigned equal weight and the overall workload is obtained by summing the values and dividing by the number of factors used. Studies have shown [24], [40], [45], [59] this modified version to be as effective as the original, with the added benefit of being a much simpler approach.

In addition, Hart and Staveland [25] determined the individual factors can be used independently to garner information about the various aspects of workload. Hart [24], in her review of the usage of NASA-TLX states the analysis of subscale ratings instead of, or in addition to, an overall rating demonstrates "one of the continuing strengths of the scale: the diagnostic value of the component subscales." Studies have also adapted the measure in various ways: using a 5-point scale [20], [54], rather than the original 21-point scale; changing the wording to increase the relevance to the tasks [26], [41] and using only some of the subscales [19], [58].

Cued Retrospective Think-Aloud (Cued RTA)

Think-aloud approaches can be Concurrent Think-Aloud (CTA), where a user articulates their thoughts as they carry out tasks on a system, or Retrospective Think-Aloud (RTA), where the user describes their thought processes after completion of the tasks [33]. This process clarifies the user's attitudes towards the system and identifies elements of the design that are problematic [30]. The think-aloud sessions are recorded and a scribe may take additional notes of the user's comments and actions [42]. For this study, CTA is an inappropriate approach to use with eye tracking, as it can cause bias to the user's first impression and affect their visual fixations [35]. When used alone, RTA can be problematic, as the participant uses memory to recall and describe their cognitive processes. This approach can result in forgotten information or retrospective justification of their actions, leading to erroneous data [2]. Cued RTA, which rather than relying on memory, requires the user's interaction to be played back to them, tends to be more effective at eliciting comments than un-cued RTA [2]. While van den Haak, de Jong and Jan Schellens [57] found RTA and CTA comparable in terms of the number of usability problems identified, using eye tracking with cued RTA allows the researcher to combine qualitative and quantitative data. It is also used in order to supplement the eye tracking data, which only shows the pattern of eye movement without any indication of why the user focused on particular parts of the screen [33]. Understanding why users focus on particular parts of the screen is important as the reasons can be positive or negative (e.g., a long fixation can reflect interest or difficulty in understanding).

3. Experiment Design

3.1. Lessons Learned from the Pilot

The key lessons ascertained from the pilot were to ensure participants: 1) are fully briefed before they commence the test; 2) understand they are not the subject of the experiment; 3) perform the interaction as instructed working with neither haste nor labouring the tasks; and 4) encourage participants to comment freely during the cued RTAs without inhibition.

Pilot participants were presented with a series of webpages corresponding to the eight variants of opt-out decision constructs. The webpages were based on real interactions encountered on B2C websites but contained only a single decision construct to minimise the impact of extraneous variables in the experiment. Based on the experience of the pilot study, several changes are to be made to the experiment. Firstly, as mentioned earlier, of the three dimensions within the eight decision constructs (i.e., default value, framing, persuasion) persuasion was found to have limited influence on participants' performance and will be eliminated. Secondly, with respect to the cued RTA, it would be impractical to conduct these sessions for every participant within the main study. Hence, a proportion of participants (e.g., 20%), will be considered a representative sample for the cued RTA.

Three types of data captured pilot participants' responses: 1) eye tracking data quantified participants' interactions with the decision constructs (e.g., duration of the interaction with the webpage; the total dwell time; the number of fixations during the interaction and the mean duration of each fixation); 2) aggregated cued RTA data explained how participants perceived their interactions during the experiment and afforded a nuanced understanding of participants' internal dialogue; and 3) heat maps (e.g., fixation intensity) provided a visual representation of participants' interaction with the webpages. This mixed method approach allowed the researchers to identify which construct variants slow down decision making and to generate a more holistic comprehension of participants' decision making and their rationale. Hence, this process of data collection and analysis will persist in the main study with the addition of the use of NASA-TLX to assess cognitive load.

3.2. The Proposed Study

This study involves both quantitative (i.e., eye tracking and NASA-TLX) and qualitative (i.e., cued RTA) data. To ensure sufficient numbers, about 100 people will participate in the eye tracking and NASA-TLX. A subset of 20 participants will be selected at random for the cued RTA. Pernice and Nielsen [47] recommend a minimum of 6 participants for qualitative eye tracking; therefore, 20 participants is more than adequate for this study.

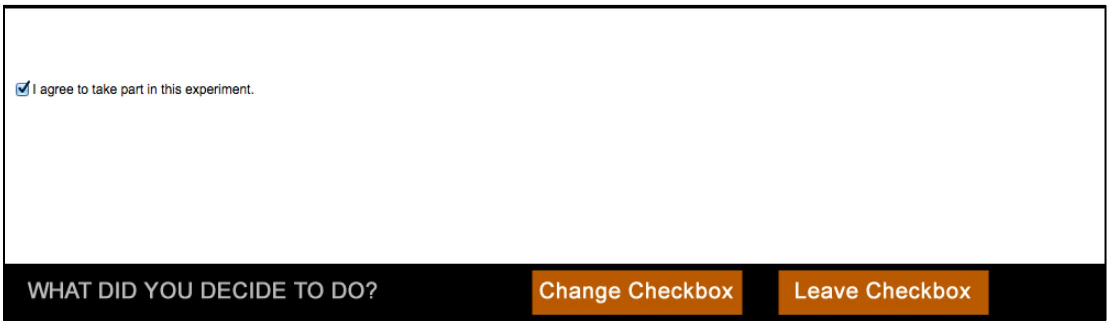

Fig. 2. Sample decision construct.

Participants will be brought into the lab and the experiment will be explained, including a description of how the equipment works, and details on the interaction with the equipment and the test material. Once this is complete, the equipment is calibrated and validated to ensure the readings are accurate. Before the experiment begins, the participant will engage with a test run to ensure they understand how to interact with the decision constructs and how to complete the NASA-TLX screens. They will initially be presented with a test screen (see Figure 2) and told which decision to make (i.e., to agree to participate in the experiment) and then click on the appropriate button (i.e., Change Checkbox or Leave Checkbox). They will then complete the NASA-TLX for the interaction with the decision construct. For this study, the NASA-TLX factors most relevant are mental demand, performance and frustration. The participant will be asked to indicate the extent to which each factor contributes to their subjective perception of physical and mental workload (see Figure 3). A 5-point scale is used.

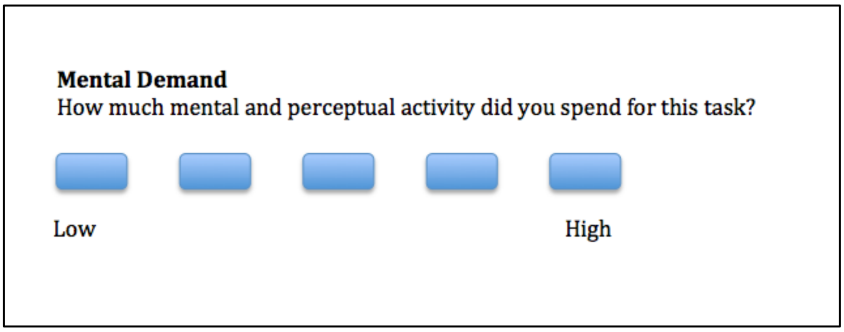

Fig. 3. NASA-TLX rating scale.

The participants will be presented with a series of opt-out decisions, consisting of both pre-selected and un-selected constructs and using different types of framing. While the constructs are based on interactions on real B2C websites, the wording is framed to ensure consistency. In addition, each decision construct will be presented independently, unlike real websites, which often contain multiple decisions on a single page. While this approach reduces the realism of the interaction, it is necessary to minimise the effect of extraneous variables on the results. The decision construct and the 'change' and 'leave' buttons are located in the same area of each webpage. A total of 4 webpages will be developed, each corresponding to one of the 4 variants of opt-outs detailed earlier in Figure 1. The order in which the pages are presented will be randomised to minimise bias. The participant are told the decision required for each screen and instructed to click on the appropriate button.

Once the eye tracking is completed, a representative subset of 20 of the participants will complete the cued RTA. They will be shown an animation of their interaction with each decision construct. While they watch the animation, the participant will be asked to articulate the thought processes and feelings they had during the interaction. The interviewer will use prompts to encourage verbalisation by the participant. These prompts will be similar to the ones used in the pilot study [4] and include:

- What do you think of the wording?
- How clear was the required action?
- How easy or difficult was it to make the decision?
- What do you think would be the default action if you do nothing?

The prompts ensure all areas of interest are explored rather than directing the conversation in a specific direction. They are most helpful when interacting with participants who are finding it difficult to articulate their thoughts [42].

3.3. Analysis of Data

The data analysis will consist of several parts (see Figure 4). The various forms of presentation format (un-selected rejection framing; pre-selected acceptance framing; pre-selected rejection framing and pre-selected neutral framing) will serve as predictor (or independent) variables in a multiple linear regression model of their relationship with each of the response (or dependent) variables: duration and error rate. Additionally, the presentation format will act as predictor variables in determining their relationship with cognitive load.

Cognitive load will be measured subjectively using the NASA-TLX and physiologically using the eye tracking data. The NASA-TLX will initially be used as a benchmark in order to determine the efficacy of the various physiological measures through the use of correlation tests. This will indicate which physiological measures can be classified as valid measures of cognitive load and further analysed. This is important, give the short duration of the user interaction as much previous research involves longer tasks. For example, the short duration of the tasks may mean blink rate is only suitable for tasks longer than 3 minutes [7]. Eye Point of Gaze (EPOG), which considers fixations, saccades and dwell time, can be used to determine cognitive load. Sensitivity is acceptable if primary parameters (e.g. fixation count) are combined with secondary parameters (e.g. dwell time) [7]. Therefore, the fixation count and the dwell time for the tasks will be compared. To analyse blink rate, dwell time and fixation count for the AOI within each webpage, one-way ANOVA can be used to identify the range of participants' physiological cognitive load.

As each of the NASA-TLX subscales can stand independently [24], they can be compared for each task. An overall workload can be calculated by combining the raw ratings to produce a composite value. According to Hart [24], the overall workload can also be assessed without weighting the individual subscales. She reports 29 studies which compared the use of unweighted (known as Raw TLX) and weighted subscales. Some reported higher sensitivity, some lower and some no difference.

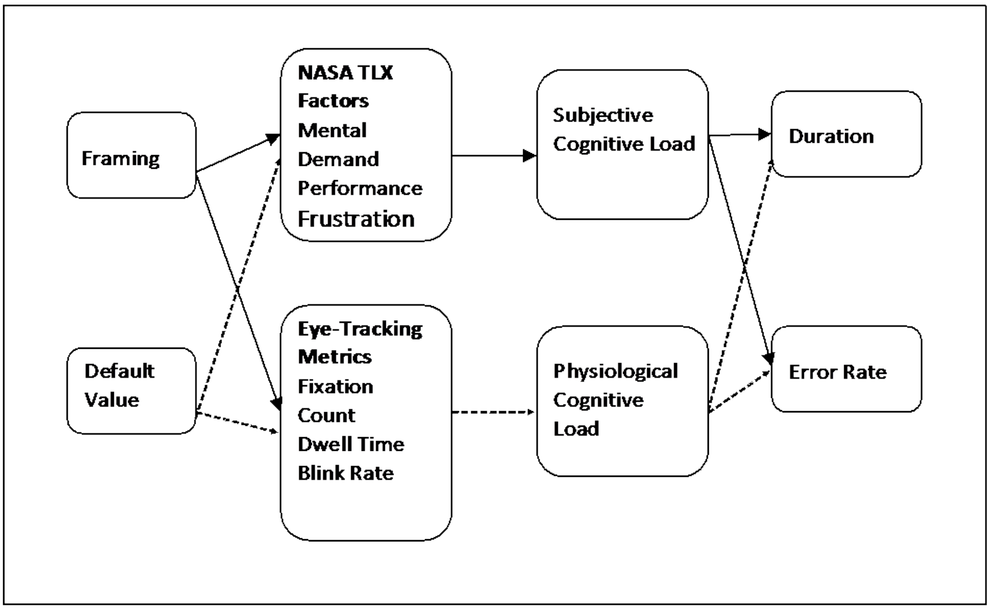

Fig. 4. Analysis of data.

4. Conclusions

This paper has traced the emergence, development and rationale of a significant study that will shed light on how participants respond to cognitive load during interactions with decision constructs on webpages. The authors have briefly described a taxonomy that identifies all decision constructs encountered during the online commercial transactional process. From these constructs, one type - the opt-out - was selected to study in more detail, as it was a construct that caused users particular difficulty and confusion. Cognitive load will be measured both physiologically and subjectively. The exploration of multiple measurements of cognitive load make it possible to gather strong quantitative data on participant interaction alongside cued RTA which capture, qualitatively, what participants will have thought and felt during the interaction. This research on decision making during webpage interactions will make an original and novel contribution to ethical design considerations of information systems development.

References

1. Albers, M.: Tapping as a Measure of Cognitive Load and Website Usability. In: 29th ACM International Conference on Design of Communication, ACM, Pisa, Italy (2011)
2. Ball, L.J., Eger, N., Stevens, R., Dodd, J.: Applying the PEEP Method in Usability Testing. Interfaces. 67, 15-19 (2006)
3. Authors: Low-Cost Carriers and High-Tech Barriers - User Views on Questionable Web Design Practices in Ireland. Irish Journal of Management. 31(1), 43-58 (2011)
4. Authors: Framing or Gaming? Constructing a Study to Explore the Impact of Option Presentation on Consumers. In: 24th International Conference on Information Systems Development, Harbin, China (2015)
5. Brunken, R., Plass, J.L., Leutner, D.: Direct Measurement of Cognitive Load in Multimedia Learning. Educational Psychologist. 38(1), 53-61 (2003)
6. Cain, B.: A Review of the Mental Workload Literature, Defence Research and Development Toronto, CANADA, (2007)

7. Castor, M., Hanson, E., Svensson, E., NäHlinder, S., Le Blaye, P., Macleod, I., Wright, N., Alfredson, J., ÅGren, L., Berggren, P., Juppet, V.R., Hilburn, B., Ohlsson, K.: Garteur Handbook of Mental Workload Measurement. Group for Aeronautical Research and Technology in EURope (GARTEUR) Flight Mechanics Action Group FM AG13 (2003)
8. Colligan, L., Potts, H.W., Finn, C.T., Sinkin, R.A.: Cognitive Workload Changes for Nurses Transitioning from a Legacy System with Paper Documentation to a Commercial Electronic Health Record. International Journal of Medical Informatics. 84(7), 469-76 (2015)
9. Day, R.F., Shyi, G.C.W., Wang, J.C.: The Effect of Flash Banners on Multiattribute Decision Making: Distractor or Source of Arousal? Psychology & Marketing. 23(5), 369-382 (2006)
10. Di Stasi, L.L., Antolí, A., Gea, M., Cañas, J.J.: A Neuroergonomic Approach to Evaluating Mental Workload in Hypermedia Interactions. International Journal of Industrial Ergonomics. 41(3), 298-304 (2011)
11. Djamasbi, S., Siegel, M., Skorinko, J., Tullis, T.: Online Viewing and Aesthetic Preferences of Generation Y and the Baby Boom Generation: Testing User Web Site Experience through Eye Tracking. International Journal of Electronic Commerce. 15(4), 121-158 (2011)
12. Djamasbi, S., Siegel, M., Tullis, T., Dai, R.: Efficiency, Trust, and Visual Appeal: Usability Testing through Eye Tracking. In: 43rd Hawaii International Conference on System Sciences (HICSS), IEEE, (2010)
13. Eveland, W.P., Dunwoody, S.: Examining Information Processing on the World Wide Web Using Think Aloud Protocols. Media Psychology. 2(3), 219-244 (2000)
14. Finomore, V.S., Shaw, T.H., Warm, J.S., Matthews, G., Boles, D.B.: Viewing the Workload of Vigilance through the Lenses of the NASA-TLX and the Mrq. Human Factors: The Journal of the Human Factors and Ergonomics Society. 55(6), 1044-1063 (2013)
15. Galy, E., Cariou, M., Melan, C.: What Is the Relationship between Mental Workload Factors and Cognitive Load Types? International Journal of Psychophysiology. 83(3), 269-75 (2012)
16. Glöckner, A., Herbold, A.K.: An Eye-Tracking Study on Information Processing in Risky Decisions: Evidence for Compensatory Strategies Based on Automatic Processes. Journal of Behavioral Decision Making. 24(1), 71-98 (2011)
17. Goh, K.N., Chen, Y.Y., Lai, F., Daud, S.C., Sivaji, A., Soo, S.-T.: A Comparison of Usability Testing Methods for an E-Commerce Website: A Case Study on a Malaysia Online Gift Shop. In: Tenth International Conference on Information Technology: New Generations (ITNG). IEEE (2013)
18. Goldberg, J.H.: Measuring Software Screen Complexity: Relating Eye Tracking, Emotional Valence, and Subjective Ratings. International Journal of Human-Computer Interaction. 30(7), 518-532 (2014)
19. González Gutiérrez, J.L., Jiménez, B.M., Hernández, E.G., López López, A.: Spanish Version of the Swedish Occupational Fatigue Inventory (Sofi): Factorial Replication, Reliability and Validity. International Journal of Industrial Ergonomics. 35(8), 737-746 (2005)
20. Goodman, J., Brewster, S.A., Gray, P.: How Can We Best Use Landmarks to Support Older People in Navigation? Behaviour & Information Technology. 24(1), 3-20 (2005)
21. Gopher, D., Braune, R.: On the Psychophysics of Workload: Why Bother with Subjective Measures? Human Factors. 26(5), 519-532 (1984)
22. Grimes, G.M., Valacich, J.S.: Mind over Mouse: The Effect of Cognitive Load on Mouse Movement In: Thirty Sixth International Conference on Information Systems, Forth Worth, USA (2015)

23. Hankins, T.C., Wilson, G.F.: A Comparison of Heart Rate, Eye Activity, Eeg and Subjective Measures of Pilot Mental Workload During Flight. Aviation, Space, and Environmental Medicine. 69(4), 360-367 (1998)
24. Hart, S.G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, (2006)
25. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology. 52, 139-183 (1988)
26. Hayashi, T., Kishi, R.: Utilization of NASA-TLX for Workload Evaluation of Gaze-Writing Systems. In: IEEE International Symposium on Multimedia, (2014)
27. Hill, S.G., Ianecchia, H.P., Byers, J.C., Bittner, A.C., Zakland, A.L., Christ, R.E.: Comparison of Four Subjective Workload Rating Scales. Human Factors. 34(4), 429-439 (1992)
28. Authors: An Eye Tracking Pilot Study on Optionality - Some Lessons Learned. In: 17th Irish Academy of Management Annual Conference, Galway (2015)
29. Hollender, N., Hofmann, C., Deneke, M., Schmitz, B.: Integrating Cognitive Load Theory and Concepts of Human–Computer Interaction. Computers in Human Behavior. 26(6), 1278-1288 (2010)
30. Holzinger, A.: Usability Engineering Methods for Software Developers. Communications of the ACM. 48(1), 71-74 (2005)
31. Hoonakker, P., Carayon, P., Gurses, A., Brown, R., Mcguire, K., Khunlertkit, A., Walker, J.M.: Measuring Workload of Icu Nurses with a Questionnaire Survey: The NASA Task Load Index (TLX). IIE transactions on healthcare systems engineering. 1(2), 131-143 (2011)
32. Huang, Y.-F., Kuo, F.-Y.: An Eye-Tracking Investigation of Internet Consumers' Decision Deliberateness. Internet Research. 21(5), 541-561 (2011)
33. Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K.-J., Lehtinen, M.: Gaze Path Stimulation in Retrospective Think-Aloud. Journal of Eye Movement Research. 2(4), 1-18 (2008)
34. Ignacio Madrid, R., Van Oostendorp, H., Puerta Melguizo, M.C.: The Effects of the Number of Links and Navigation Support on Cognitive Load and Learning with Hypertext: The Mediating Role of Reading Order. Computers in Human Behavior. 25(1), 66-75 (2009)
35. Kim, B., Dong, Y., Kim, S., Lee, K.-P.: Development of Integrated Analysis System and Tool of Perception, Recognition, and Behavior for Web Usability Test: With Emphasis on Eye-Tracking, Mouse-Tracking, and Retrospective Think Aloud, in Usability and Internationalization. Hci and Culture. Springer: Berlin Heidelberg. pp. 113-121 (2007)
36. Lee, J., Ahn, J.-H.: Attention to Banner Ads and Their Effectiveness an Eye Tracking Approach. International Journal of Electronic Commerce. 17(1), 119-137 (2012)
37. Longo, L., Rusconi, F., Noce, L., Barrett, S.: The Importance of Human Mental Workload in Web Design. In: WEBIST2012 8th International Conference on Web Information Systems and Technologies, (2012)
38. Luximon, A., Goonetilleke, R.S.: Simplified Subjective Workload Assessment Technique. Ergonomics. 44(3), 229-43 (2001)
39. Martin, S.: Measuring Cognitive Load and Cognition: Metrics for Technology-Enhanced Learning. Educational Research and Evaluation. 20(7-8), 592-621 (2014)
40. Miller, S.: Literature Review - Workload Measures. National Advanced Driving Simulator (2001)
41. Mohamed, R., Raman, M., Anderson, J., Mclaughlin, K., Rostom, A., Coderre, S.: Validation of the National Aeronautics and Space Administration-Task Load Index as a Tool to Evaluate the Learning Curve for Endoscopy Training. Canadian Journal of Gastroentrology & Hepatology. 28(3), 155-160 (2014)
42. Monk, A., Davenport, L., Haber, J., Wright, P.: Improving Your Human-Computer Interface: A Practical Technique. Prentice Hall, New York, USA (1993)

43. Nasa: NASA-TLX Paper and Pen Manual. (1986); Available from: http://humansystems.arc.nasa.gov/groups/tlx/paperpencil.html
44. Nielsen, J., Pernice, K.: Eyetracking Web Usability. New Riders, Berkeley, CA (2010)
45. Nygren, T.E.: Psychometric Properties of Subjective Workload Measurement Techniques: Implications for Their Use in the Assessment of Perceived Mental Workload. Human Factors. 33(1), 17-33 (1991)
46. Paas, F.: Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. Journal of Educational Psychology. 84(4), 429-434 (1992)
47. Pernice, K., Nielsen, J.: Eyetracking Methodology: How to Conduct and Evaluate Usability Studies Using Eyetracking. Nielsen Norman Group Technical Report. (2009)
48. Resnick, M., Albert, W.: The Impact of Advertising Location and User Task on the Emergence of Banner Ad Blindness: An Eye-Tracking Study. International Journal of Human-Computer Interaction. 30(3), 206-219 (2014)
49. Rubio, S., Díaz, J.E., Martín, J.S., Puente, J.M.: Evaluation of Subjective Mental Workload: A Comparison of Swat, NASA-TLX, and Workload Profile Methods. Applied Psychology: An International Review. 53(1), 61-86 (2004)
50. Salmon, P., Stanton, N., Baber, C., Walker , G., Green, D.: Human Factors Design & Evaluation Methods Review, (2004)
51. Schmutz, P., Heinz, S., Métrailler, Y., Opwis, K.: Cognitive Load in Ecommerce Applications—Measurement and Effects on User Satisfaction. Advances in Human-Computer Interaction. 1-9 (2009)
52. Sendurur, E., Yildirim, Z.: Students' Web Search Strategies with Different Task Types: An Eye-Tracking Study. International Journal of Human-Computer Interaction. 31(2), 101-111 (2015)
53. Sivaji, A., Downe, A.G., Mazlan, M.F., Soo, S.-T., Abdullah, A.: Importance of Incorporating Fundamental Usability with Social & Trust Elements for E-Commerce Website. In: International Conference on Business, Engineering and Industrial Applications (ICBEIA), IEEE, (2011)
54. Smuts, M., Scholtz, B., Calitz, A.P.: Usability Guidelines for Designing Information Visualisation Tools for Novice Users. In: Beyond development. Time for a new ICT4D paradigm? Proceedings of the 9th IDIA conference, IDIA2015, Nungwi, Zanzibar (2015)
55. Szulewski, A., Roth, N., Howes, D.: The Use of Task-Evoked Pupillary Response as an Objective Measure of Cognitive Load in Novices and Trained Physicians: A New Tool for the Assessment of Expertise. Academic Medicine. 90(7), 981-7 (2015)
56. Authors: Opaque Web Practices among Low-Cost Carriers. Journal of Air Transport Management. 15, 299-307 (2009)
57. Van Den Haak, M., De Jong, M., Jan Schellens, P.: Retrospective Vs. Concurrent Think-Aloud Protocols: Testing the Usability of an Online Library Catalogue. Behaviour & Information Technology. 22(5), 339-351 (2003)
58. Wesiaka, G., Steinera, C.M., Moore, A., Dagger, D., Power, D., Bertholda, M., Alberta, D., Conlan, O.: Iterative Augmentation of a Medical Training Simulator: Effects of Affective Metacognitive Scaffolding1. Computers & Education 76, 13-19 (2014)
59. Wiebe, E.N., Roberts, E., Behrend, T.S.: An Examination of Two Mental Workload Measurement Approaches to Understanding Multimedia Learning. Computers in Human Behavior. 26(3), 474-481 (2010)
60. Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O.N., Atkins, M.S.: Workload Assessment of Surgeons: Correlation between NASA-TLX and Blinks. Surgical Endoscopy. 26(10), 2746-50 (2012)